Title

Sex-Dimorphic Analyses Identify Novel and Sex-Specific Genetic Associations in Inflammatory Bowel Disease.

Journal

Inflammatory Bowel Diseases, 29(10)

Authors

Khrom, Michelle
Li, Dalin
Naito, Takeo
et al.

Publication Date

2023-10-03

DOI

10.1093/ibd/izad089

Peer reviewed

Inflammatory Bowel Diseases, 2023, **29**, 1622–1632
https://doi.org/10.1093/ibd/izad089
Advance access publication 1 June 2023
Original Research Article - Basic Science

Sex-Dimorphic Analyses Identify Novel and Sex-Specific Genetic Associations in Inflammatory Bowel Disease

Michelle Khrom, BS,[*] Dalin Li, PhD,[*] Takeo Naito, MD, PhD,[*] Ho-Su Lee, MD, PhD,[†] Gregory J. Botwin, BS,[*] Alka A. Potdar, PhD,[*] Gabrielle Boucher, PhD,[§] NIDDK Inflammatory Bowel Disease Genetics Consortium, International Inflammatory Bowel Disease Genetics Consortium, Shaohong Yang, MD,[*] Emebet Mengesha, BS,[*] Shishir Dube, PhD,[*] Kyuyoung Song, PhD,[†] Dermot P.B. McGovern, MD, PhD,[*] and Talin Haritunians, PhD[*]

From the
[*]F. Widjaja Inflammatory Bowel Disease Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA;
[†]Department of Biochemistry and Molecular Biology, University of Ulsan College of Medicine, Seoul, Korea; and
[§]Research Center, Montreal Heart Institute, Montreal, Quebec, Canada

Address correspondence to: Talin Haritunians, PhD, Cedars-Sinai Medical Center, 8700 Beverly Boulevard, Davis 4094B, Los Angeles, CA 90048, USA (talin.haritunians@cshs.org).

Abstract

Background: Sex is an integral variable often overlooked in complex disease genetics. Differences between sexes have been reported in natural history, disease complications, and age of onset in inflammatory bowel disease (IBD). While association studies have identified >230 IBD loci, there have been a limited number of studies investigating sex differences underlying these genetic associations.

Methods: We report the first investigation of sex-dimorphic associations via meta-analysis of a sex-stratified association study (34 579 IBD cases, 39 125 controls). In addition, we performed chromosome (chr) X–specific analyses, considering models of X inactivation (XCI) and XCI escape. Demographic and clinical characteristics were also compared between sexes.

Results: We identified significant differences between sexes for disease location and perianal complication in Crohn's disease and disease extent in ulcerative colitis. We observed genome-wide-significant sex-dimorphic associations ($P < 5 \times 10^{-8}$) at loci not previously reported in large-scale IBD genetic studies, including at chr9q22, *CARMIL1*, and *UBASH3A*. We identified variants in known IBD loci, including in chr2p15 and within the major histocompatibility complex on chr6, exhibiting sex-specific patterns of association ($P < 5 \times 10^{-7}$ in one sex only). We identified 3 chrX associations with IBD, including a novel Crohn's disease susceptibility locus at Xp22.

Conclusions: These analyses identified novel IBD loci, in addition to characterizing sex-specific patterns of associations underlying sex-dimorphic associations. By elucidating the role of sex in IBD genetics, our study will help enhance our understanding of the differences between the sexes in IBD biology and underscores a need to move beyond conventional sex-combined analyses to appreciate the genetic architecture of IBD more comprehensively.

Lay Summary

Sex-dimorphic meta-analyses of sex-stratified case-control (n = 73 704) regression identified 3 novel inflammatory bowel disease loci reaching genome-wide significance and highlighted chromosome 2 and major histocompatibility complex variants exhibiting sex-specific association. In addition, a novel chromosome X Crohn's disease susceptibility locus was identified.

Key Words: sex dimorphic, genetics, inflammatory bowel disease

Introduction

Sex is an important biological factor often overlooked in genome-wide association studies of complex traits. Sex differences in disease prevalence, severity, and natural history have been documented in cardiovascular disease, psychiatric disorders, Alzheimer's, Parkinson's, and numerous immune-mediated diseases.[1] Inflammatory bowel disease (IBD), including Crohn's disease (CD) and ulcerative colitis (UC), is a complex disease of immune dysregulation to commensal gut flora causing chronic inflammation of the gastrointestinal (GI) tract in genetically susceptible individuals. In IBD, a higher prevalence in females has been observed in CD.[2] A recent meta-analysis of 17 population-based cohorts demonstrated sex differences in IBD age of onset.[3] In addition, it has also been reported that males diagnosed with IBD are more likely to present with upper GI involvement, have a higher likelihood of ileal CD diagnosis, and undergo more small bowel surgery, while colonic disease and extraintestinal manifestations are more common in females.[4,5]

Received for publication: May 11, 2022. Editorial Decision: April 12, 2023

Key Messages:

- Studies have implicated >230 inflammatory bowel disease (IBD) susceptibility loci predominantly through sex-combined genetic association analyses largely focused on autosomes.
- This is the largest study of IBD phenotypes across the sexes and the first investigation of the role of sex underlying IBD genetic associations utilizing sex-dimorphic meta-analysis of a sex-specific association study, in addition to analysis of variants on chromosome X.
- Precision medicine approaches in IBD are needed to improve clinical outcomes; these advances require both an understanding of clinical variability and the differences in underlying molecular signatures in diverse populations including the sexes.

Differences between the female and male immune systems are well established. Males have higher natural killer cell numbers, while macrophage and neutrophil phagocytic activity and the efficiency of antigen-presenting cells have been reported to be higher in females. Females also have a higher frequency of $CD4^+$ T cells, higher CD4-to-CD8 ratio, higher B cell numbers, and greater antibody production, and generally produce a T helper 2–dominant response, whereas males have a higher number of $CD8^+$ T cells and T regulatory (T-reg) cells and an observed T helper 1 cell bias.[6] Sex steroid hormones and hormonal changes throughout puberty, adulthood, and especially pregnancy contribute greatly to these differences in immunity. For example, progesterone and testosterone are generally anti-inflammatory, while estradiol (E2) has been shown to exert either proinflammatory or anti-inflammatory effects depending on concentrations (low vs high E2 concentrations, respectively).[6]

In addition to the differences noted previously, sex differences in genetic variant association, gene regulation, or gene expression likely contribute to the sexual dimorphism of complex diseases.[7,8] Most sex-specific genetic associations in IBD have been observed in autosomal loci,[9-11] and chromosome X (chrX) has traditionally been excluded due to a previous lack of statistical tools to properly account for chrX inactivation (XCI) and dosage compensation between the sexes. Approximately 15% of chrX genes escape XCI to varying degrees. This phenomenon of incomplete and variable XCI can be very tissue specific, resulting in sex-biased gene expression that may potentially contribute to sex-specific phenotype diversity.[12] Across all traits, only ~550 associations have been reported on chrX through genome-wide association studies (GWASs). Given that chrX carries 848 protein-coding genes, including numerous immune response–related genes, it is perceivable that a more thorough investigation will highlight a growing role for chrX genes in immune-mediated complex diseases.[13]

More than 230 IBD susceptibility loci have been identified thus far,[14,15] yet sex differences underlying these associations have not been thoroughly investigated, and the role of sex chromosomes in IBD has been largely ignored, with a few exceptions. In this study, we utilized software tools specific for sex-dimorphic analyses of autosomal variants and the analysis of chrX variants[16,17] to identify novel IBD associations at both autosomal and chrX loci, in addition to highlighting sex-specific patterns of association underlying observed sex-dimorphic associations. Our results demonstrate that investigating sex-specific genetic architecture by moving beyond more conventional sex-combined analyses as well as incorporating sex chromosomes in association studies may yield interesting biological insight and contribute toward better addressing the issue of missing heritability in IBD.

Methods

Study subjects

A total of 5798 IBD cases and 7738 non-IBD controls were available from Cedars-Sinai Medical Center (CSMC). A subset of non-IBD controls included the British Birth Cohort 1958 and samples provided by the National Laboratory for the Genetics of Israeli Populations (https://www.tau.ac.il/medicine/NLGIP/). Diagnosis of IBD was based on standard endoscopic, histologic, and radiographic features. Clinical phenotypes were made available through the CSMC MIRIAD Biobank. Subject recruitment in the International IBD Genetics Consortium (IIBDGC) has been documented elsewhere.[14] Briefly, subjects were recruited from 15 countries in Europe, North America, and Oceania. Diagnosis of IBD was based on accepted radiological, endoscopic, and histopathological evaluation. A total of 30 779 IBD cases and 33 778 non-IBD controls were available from IIBDGC after exclusion of any overlapping CSMC subjects.

Genotyping and quality control

All subjects were genotyped using the Illumina Infinium Immunochip-v1 array per manufacturer's protocol (Illumina). For CSMC (including National Laboratory for the Genetics of Israeli Populations and British Birth Cohort 1958), average genotyping rate and replicate concordance for samples that passed quality control (QC) were 99.80% and 99.99%, respectively. Samples of non-European ethnicity (as defined by <70% European admixture) were excluded from analyses. Samples were excluded if demonstrating sex inconsistencies, missingness >3%, or a high proportion of identity-by-descent (Pi-hat > 0.1875). IIBDGC Immunochip-v1 array genotyping and European ancestry sample and variant QC has been previously described.[14] Post-genotyping single nucleotide polymorphism (SNP) QC for both cohorts excluded variants with missingness >3%, minor allele frequency <3%, deviation from Hardy-Weinberg equilibrium ($P_{controls} < 1 \times 10^{-4}$), and differential missingness between sexes or between cases and controls ($P < 1 \times 10^{-4}$). ChrX-specific QC metrics included exclusion of variants with differential minor allele frequency and missingness ($P < 1 \times 10^{-4}$) between male and female controls, in addition to deviation from Hardy-Weinberg equilibrium in females ($P_{F\text{-}controls} < 1 \times 10^{-4}$). A total of 9917 and 63 787 samples and 97 261 and 101 109 SNPs from CSMC and IIBDGC, respectively, were available post-QC. QC was performed using PLINK[18] and R v4.2.0 (R Foundation for Statistical Computing, https://www.R-project.org/), and XWAS[16] for chrX.

Statistical analysis

Univariate analyses of clinical and demographic variables were performed in R using the chi-square test or regression for categorical or continuous variables, respectively. Single SNP association of autosomal variants was performed separately for males and females using logistic (for case-control

disease status and categorical clinical variables) or linear (for age-related variables) regression, with adjustment for population substratification with principal components using PLINK,[18] followed by inverse variance–weighted fixed-effects meta-analyses using GWAMAv2.2.2.[17] Three meta-analyses strategies were implemented in GWAMA: (1) sex-specific, (2) sex-combined, and (3) sex-dimorphic, which represents a combined *P* value of male- and female-specific estimates while allowing for heterogeneity in effects between sexes (chi-square distribution with 2 degrees of freedom [df]). These 2-df sex-dimorphic meta-analyses are, in theory, equivalent to testing for association allowing for interaction between SNP and sex under an additive model. In addition, GWAMA reports a 1-df Cochran's Q test for heterogeneity between sexes. See **Figure 1** for analytic schematic for Immunochip-wide autosomal variants.

Linkage disequilibrium (LD)–based clumping for autosomal variants within 500 kb and r^2 >0.5 was performed using PLINK to identify index SNPs. LD for the CEU population was also evaluated using LDlink.[19] SNPs in LD r^2 ≥0.5 with lead variants were excluded, unless otherwise noted. For previously reported IBD associated variants, proxies r^2 ≥0.8 (CEU) were used if reported variant failed Immunochip genotyping QC.[15,19] Uncontrolled confounding from population substructure for the combined CSMC-IIBDGC cohort was evaluated by calculating the genomic control inflation factor for 1000 random cases and 1000 random controls ($\lambda_{GC\text{-}1000}$) using a set of null SNPs that passed QC and have not been associated with immune-mediated diseases.[14,20] Test statistics showed negligible inflation (**Supplementary Table 1**).

Given the unique nature of chrX, we performed chrX association analysis under the models of (1) complete and uniform XCI (XCI-Complete) in females, with hemizygous males corresponding to homozygous females and a similar effect size between sexes (PLINK regression with --xchr-model 2, with sex as a covariate and male chrX genotypes coded as 0 or 2, female chrX genotypes coded as 0, 1, or 2) and (2) escape from XCI (XCI-Escape) in females, as some genes are known to escape X inactivation and others shown to have variable, tissue-specific escape (--xchr-model 1, with sex as a covariate and males coded as 0 or 1).[12] Because assumptions of complete XCI and equal effects sizes do not always hold, we also performed a third model, of sex-stratified regression analyses separately in males and females and combined by a weighted z-score approach for Fisher's test (Fisher combined, allowing for differential effects size and direction of effect between sexes and unaffected by allele coding in males) or weighted Stouffer's method (Stouffer combined, accounting for different sample sizes and direction of effect) (XWAS).[16]

Genes names cited throughout the manuscript and in tables refer to gene(s) nearest the associated variant of interest, with the exception of previously reported IBD variants in which both the gene and implicated gene columns are as reported in de Lange et al.[15] Variants were annotated as novel if they fell outside 1 Mb of previously reported variant(s) in an IBD-associated genomic region and r^2 <0.2 with previously reported SNPs.[15] RegulomeDBv2.0, which integrates data from ENCODE and Roadmap, was utilized to investigate DNA features and regulatory elements overlapping associated SNPs.[21] Expression quantitative trait loci (eQTLs) and splicing quantitative trait loci were examined using data available from GTEx and the eQTLGen Consortium.[22,23]

Results

Study population

Cohort demographic and clinical characteristics are shown in **Table 1**. For males, there was a total 8653 CD, 7527 UC, 241 IBD–unclassified, and 16 780 controls; for females, there was a total of 10 760 CD, 7196 UC, 202 IBD–unclassified, and 22 345 controls. Females had a higher mean age at diagnosis at CSMC (26.46 years vs 25.89 years), while in the IIBDGC, males had a higher mean age at diagnosis (30.66 years vs 31.97 years). Disease duration was comparable between sexes in CSMC, whereas there was a slightly longer mean disease duration in females vs males in IIBDGC (23.76 years and 22.94 years). In both the CSMC and IIBDGC cohorts, the most frequently observed phenotypes in CD subjects were inflammatory (non-stricturing, non-internal penetrating) disease behavior (B1) and ileocolonic disease location (L3). No significant differences in disease behavior were observed between sexes for either cohort. Males were less likely to have a colonic-only disease location (L2) ($P = .045$ and 4.3×10^{-4} for CSMC and IIBDGC, respectively). Males were more likely to have upper GI disease (IIBDGC only: $P = 7.7 \times 10^{-8}$) and perianal CD ($P = .022$ and .015 for CSMC and IIBDGC, respectively). UC subjects predominantly had extensive disease (E3), and males were more likely to have extensive UC than females ($P = 3.3 \times 10^{-3}$ and 2.8×10^{-8} for CSMC and IIBDGC, respectively).

Sex-dimorphic analyses identify autosomal associations in novel IBD loci

Compared with sex-combined meta-analyses, the 2-df sex-dimorphic test has been shown to be more powerful in the presence of potential heterogeneity between the sexes.[17,24] Thus, we first investigated whether the 2-df sex-dimorphic meta-analysis of sex-stratified case-control regression on 34 579 cases and 39 125 controls could reveal associations in

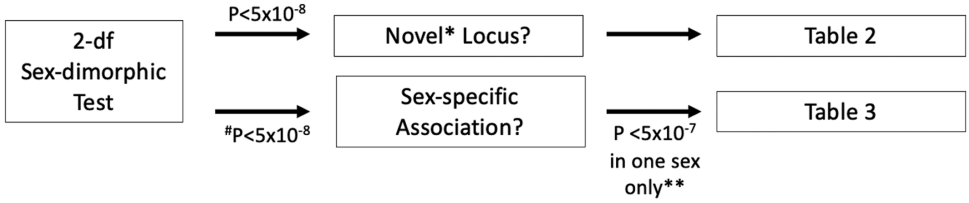

Figure 1. Schematic describing analysis workflow. *Single nucleotide polymorphism located outside the 1 Mb window of previously reported variants in an inflammatory bowel disease–associated genomic region and $r^2 < 0.2$ with previously reported single nucleotide polymorphism.[15] **$P \geq .05$ in the opposite sex in sex-specific meta-analyses. #Variants $5 \times 10^{-8} < P_{2\text{-df sex-dimorphic}} < 5 \times 10^{-7}$ evaluated for sex-specific association are listed in Supplementary Table 3.

Table 1. Cohort demographics and clinical characteristics.

	CSMC (total n = 9917)			IIBDGC (total n = 63 787)		
	Male (n = 4933)	Female (n = 4984)	*P* [OR (95%CI)]	Male (n = 28 268)	Female (n = 35 519)	*P* [OR (95%CI)]
Diagnosis						
CD	1426 (28.9)	1266 (25.4)		7227 (25.6)	9494 (26.7)	
UC	786 (15.9)	748 (15.0)		6741 (23.8)	6448 (18.2)	
IBD-U	89 (1.8)	80 (1.6)		152 (0.5)	122 (0.3)	
Control subjects	2632 (53.4)	2890 (58.0)		14 148 (50.0)	19 455 (54.8)	
Age at diagnosis, y[a]	25.89 ± 15.1	26.46 ± 14.2	2.1×10^{-3} [0.996 (0.994-0.999)][a]	31.97 ± 15.5	30.66 ± 14.5	5.0×10^{-11} [0.994 (0.993-0.996][a]
Disease duration, y[a]	21.0 ± 11.9	21.6 ± 12.5	.85 [1.0 (1.0-1.0)][a]	22.94 ± 10.6	23.76 ± 10.8	7.7×10^{-9} [1.007 (1.005-1.01)][a]
CD disease location						
Ileal only (L1)	161 (11.3)	145 (11.5)		1783 (24.7)	2213 (23.3)	
Colonic only (L2)	139 (9.7)	147 (11.6)	.045 [0.77 (0.60-1.0)][b]	1200 (16.6)	1774 (18.7)	4.3×10^{-4} [0.86 (0.79-0.94)][b]
Ileocolonic (L3)	634 (44.5)	504 (39.8)		2360 (32.7)	3056 (32.2)	
Missing	492 (34.5)	470 (37.1)		1884 (26.1)	2451 (25.8)	
CD upper GI disease						
Yes	133 (9.3)	111 (8.8)	.84 [1.03 (0.78-1.35)]	554 (7.7)	515 (5.4)	7.7×10^{-8} [1.42 (1.25-1.61)]
No	823 (57.7)	706 (55.8)		3851 (53.3)	5066 (53.4)	
Missing	470 (33.0)	449 (35.5)		2822 (39.0)	3913 (41.2)	
CD disease behavior						
Inflammatory (B1)	442 (31.0)	381 (30.1)	.87 [0.98 (0.82-1.18)][c]	2436 (33.7)	3237 (34.1)	.89 [1.01 (0.94-1.08)][c]
Stricturing (B2)	244 (17.1)	221 (17.5)		1344 (18.6)	1654 (17.4)	
Penetrating (B3)	270 (18.9)	215 (17.0)		1421 (19.7)	2039 (21.5)	
Missing	470 (33.0)	449 (35.5)		2026 (28.0)	2564 (27.0)	
CD perianal disease						
Yes	370 (25.9)	280 (22.1)	.02 [1.23 (1.03-1.48)]	1364 (18.9)	1575 (16.6)	.015 [1.11 (1.02-1.21)]
No	1056 (74.1)	983 (77.6)		3342 (46.2)	4293 (45.2)	
Missing	-	3 (0.2)		2521 (34.9)	3626 (38.2)	
UC disease extent						
Proctitis (E1)	20 (2.5)	22 (2.9)		501 (7.4)	743 (11.5)	
Left-sided (E2)	131 (16.7)	156 (20.9)		1477 (21.9)	1383 (21.4)	
Extensive (E3)	379 (48.2)	301 (40.2)	3.3×10^{-3} [1.48 (1.13-1.95)][d]	1691 (25.1)	1394 (21.6)	2.8×10^{-8} [1.30 (1.19-1.43)][d]
Missing	256 (32.6)	269 (36.0)		3072 (45.6)	2928 (45.4)	

Values are n (%) or mean ± SD.
Abbreviations: CD, Crohn's disease; CI, confidence interval; CSMC, Cedars-Sinai Medical Center; GI, gastrointestinal; IIBDGC, International IBD Genetics Consortium; IBD-U, inflammatory bowel disease unclassified; OR, odds ratio; UC, ulcerative colitis.
[a]Variables include high levels of missing data.
[b]Colonic-only disease (L2 vs L1, L3).
[c]Inflammatory disease (B1 vs B2, B3).
[d]Extensive disease (E3 vs E1, E2).

Table 2: Novel loci identified by sex-dimorphic analyses.

Trait	SNP[a]	Chromosome	Position (hg19)	Nearest gene(s)	Effect allele	Other allele	Male specific					Female specific					2-df sex-dimorphic	1-df sex-heterogeneity	Sex combined				
							OR	Lower 95% CI	Upper 95% CI	*P*	n	OR	Lower 95% CI	Upper 95% CI	*P*	n	*P*	*P*	OR	Lower 95% CI	Upper 95% CI	*P*	n
IBD	rs11788118	9	102337331	*NAMA*	A	G	0.937	0.902	0.974	1.01×10^{-3}	33198	0.901	0.871	0.933	3.92×10^{-9}	40500	1.30×10^{-10}	1.40×10^{-1}	0.917	0.894	0.941	4.83×10^{-11}	73698
IBD	rs7752195[a]	6	25419094	*CARMIL1*	A	G	0.909	0.853	0.968	2.87×10^{-3}	33195	0.854	0.806	0.905	1.19×10^{-7}	40496	9.38×10^{-9}	1.60×10^{-1}	0.879	0.842	0.917	3.44×10^{-9}	73691
IBD	rs2690110[a]	6	25328567	*CARMIL1*	G	A	1.089	1.054	1.125	3.75×10^{-7}	33194	1.051	1.021	1.083	9.65×10^{-4}	40498	1.04×10^{-8}	1.20×10^{-1}	1.068	1.045	1.092	4.76×10^{-9}	73692
UC	rs1893592	21	43855067	*UBASH3A*	C	A	0.900	0.861	0.942	4.70×10^{-6}	24307	0.923	0.883	0.964	3.22×10^{-4}	29538	4.28×10^{-8}	4.47×10^{-1}	0.912	0.883	0.941	7.88×10^{-9}	53845

Abbreviations: CI, confidence interval; IBD, inflammatory bowel disease; OR, odds ratio; SNP, single nucleotide polymorphism; UC, ulcerative colitis.
[a]Linkage disequilibrium r^2 CEU = 0.032.

novel loci not previously reported in recent large-scale IBD GWAS (**Figure 1**). Four variants in 3 loci exhibited genome-wide significance ($P < 5 \times 10^{-8}$) with 2-df sex-dimorphic meta-analyses, including with IBD, rs11788118 (chr9q22.33; $P_{\text{2df-sex-dimorphic}} = 1.3 \times 10^{-10}$) and 2 SNPs ($r^2 = 0.032$) in *CARMIL1* rs7752195 ($P_{\text{2df-sex-dimorphic}} = 9.4 \times 10^{-9}$) and rs2690110 ($P_{\text{2df-sex-dimorphic}} = 1.0 \times 10^{-8}$), and rs1893592 in *UBASH3A* and UC ($P_{\text{2df-sex-dimorphic}} = 4.3 \times 10^{-8}$) (**Table 2**). While the strength of association varied between sexes, these variants did not demonstrate distinctive sex-specific associations observed in one sex only and exhibited comparable effect sizes, and thus also demonstrated genome-wide significance in the sex-combined meta-analysis. Furthermore, no significance was observed with 1-df test for sex heterogeneity ($P_{\text{1df-sex-heterogeneity}} > .1$).

Genome-wide sex-dimorphic IBD associations demonstrate sex-specific patterns of association

Next, we interrogated the potential role of sex underlying associations demonstrating genome-wide significance with the 2-df sex-dimorphic test. Here, variants $P_{\text{2df-sex-dimorphic}} < 5 \times 10^{-8}$ were classified as sex specific when achieving significance $P < 5 \times 10^{-7}$ in one sex only ($P \geq .05$ in opposite sex) in the sex-specific meta-analyses (**Figure 1**). Thirteen, 5, and 12 variants with sex-dimorphic genome-wide significance for CD, UC, and IBD, respectively, displayed sex-specific patterns of association (**Table 3**). We observed the strongest associations with rs778160/*USP34* in CD ($P_{\text{2df-sex-dimorphic}} = 8.2 \times 10^{-15}$; $P_{\text{Female}} = 4.5 \times 10^{-15}$), rs35812145/*C2orf74* in CD and IBD ($P_{\text{2df-sex-dimorphic}} = 3.1 \times 10^{-13}$; $P_{\text{Female}} = 5.4 \times 10^{-14}$; and $P_{\text{2df-sex-dimorphic}} = 3.0 \times 10^{-12}$; $P_{\text{Female}} = 1.7 \times 10^{-12}$, respectively), rs1150755/*TNXB* in IBD ($P_{\text{2df-sex-dimorphic}} = 4.1 \times 10^{-13}$; $P_{\text{Female}} = 2.1 \times 10^{-13}$), and rs2269426/*TNXB* in UC ($P_{\text{2df-sex-dimorphic}} = 1.0 \times 10^{-10}$; $P_{\text{Male}} = 3.6 \times 10^{-11}$). In addition to the male-specific association with rs2269426/*TNXB* in UC, only 2 other male-specific associations were observed at rs72834724 in UC ($P_{\text{Male}} = 3.7 \times 10^{-9}$) and rs1060970 in CD ($P_{\text{Male}} = 2.7 \times 10^{-8}$) (**Table 3**). Variants demonstrated nominal levels of significance with 1-df test for sex heterogeneity and a third were located on chr6 within the major histocompatibility complex (MHC) (**Table 3**), interesting given the well-established role of the MHC and HLA locus in IBD.[25] We interrogated if any these variants exhibited sex-dimorphic genome-wide significance with any clinical variables. Only rs2239805/*HLA-DRA*, which demonstrated a female-specific association in UC (Table 3B), also demonstrated significance with the sex-dimorphic test with extensive UC ($P_{\text{2df-sex-dimorphic}} = 3.2 \times 10^{-8}$); however, sex-specific patterns were not evident, as the association was observed in both sexes with similar directions and magnitude of effect (**Supplementary Table 2A**). These results emphasize a need to move beyond conventional sex-combined analyses and investigate sex-dimorphic and sex-specific analyses to appreciate the genetic architecture of IBD more comprehensively.

Additional variants were identified reaching nominal levels of significance ($5 \times 10^{-8} < P < 5 \times 10^{-7}$) with 2-df sex-dimorphic meta-analyses and demonstrating sex-specific association, including associations in known loci such as rs12931474/*RMI2* (IBD: $P_{\text{2df-sex-dimorphic}} = 6.0 \times 10^{-8}$; $P_{\text{Male}} = 2.4 \times 10^{-8}$) and putative novel loci such as rs57275892/*PSMA6* ($P_{\text{2df-sex-dimorphic}} = 2.6 \times 10^{-7}$; $P_{\text{Female}} = 2.1 \times 10^{-7}$) and rs980263457/*OS9* ($P_{\text{2df-sex-dimorphic}} = 1.5 \times 10^{-7}$; $P_{\text{Male}} = 1.3 \times 10^{-7}$), as well as

Table 3: Genome-wide-significant sex-dimorphic associations for CD, UC, and IBD autosomes.

SNP[a]	Chromosome	Position (hg19)	Nearest gene(s)	Effect allele	Other allele	Male specific					Female specific					2-df sex-dimorphic *P*	1-df sex-heterogeneity *P*	Sex combined				
						OR	Lower 95% CI	Upper 95% CI	*P*	n	OR	Lower 95% CI	Upper 95% CI	*P*	n			OR	Lower 95% CI	Upper 95% CI	*P*	n
CD autosomes																						
rs35812145	2	61391300	*C2orf74*	G	A	1.024	0.977	1.073	3.29×10^{-1}	25432	1.169	1.122	1.217	5.40×10^{-14}	33097	3.08×10^{-13}	2.79×10^{-5}	1.104	1.071	1.139	2.57×10^{-10}	58529
rs6545846	2	61442910	*USP34*	A	G	1.031	0.992	1.072	1.21×10^{-1}	25426	1.110	1.073	1.148	1.58×10^{-9}	33102	3.58×10^{-9}	5.04×10^{-3}	1.075	1.048	1.103	2.61×10^{-8}	58528
rs778160	2	61557164	*USP34*	G	A	1.038	0.997	1.081	6.80×10^{-2}	25432	1.150	1.111	1.191	4.53×10^{-15}	33102	8.16×10^{-15}	1.62×10^{-4}	1.101	1.072	1.130	1.16×10^{-12}	58534
rs35518316	2	65683672	*SPRED2*	A	G	0.969	0.918	1.023	2.57×10^{-1}	25429	0.867	0.827	0.909	4.10×10^{-9}	33101	1.59×10^{-8}	2.43×10^{-3}	0.910	0.878	0.943	2.41×10^{-7}	58530
rs4676410	2	241563739	*GPR35*	A	G	1.030	0.982	1.081	2.19×10^{-1}	25430	1.128	1.082	1.175	1.20×10^{-8}	33090	4.04×10^{-8}	5.11×10^{-3}	1.085	1.052	1.119	3.13×10^{-7}	58520
rs1060970	3	49708769	*BSN*	G	A	0.875	0.835	0.917	2.66×10^{-8}	25430	0.961	0.923	1.001	5.41×10^{-2}	33105	2.92×10^{-8}	3.03×10^{-3}	0.923	0.896	0.952	3.66×10^{-7}	58535
rs9391701	6	30983263	*MUC22*	A	G	1.043	0.986	1.104	1.40×10^{-1}	25432	1.154	1.099	1.212	9.71×10^{-9}	33099	2.36×10^{-8}	8.10×10^{-3}	1.105	1.065	1.147	1.17×10^{-7}	58531
rs361525	6	31543101	*TNF*	A	G	1.076	0.990	1.171	8.50×10^{-2}	25431	1.257	1.171	1.350	3.09×10^{-10}	33105	5.49×10^{-10}	5.69×10^{-3}	1.178	1.116	1.244	3.40×10^{-9}	58536
rs12722522	10	6078553	*IL2RA*	A	G	1.052	0.990	1.118	1.03×10^{-1}	25432	1.168	1.108	1.231	7.97×10^{-9}	33102	1.53×10^{-8}	1.11×10^{-2}	1.117	1.073	1.162	5.59×10^{-8}	58534
rs17135045	11	76325282	*EMSY;LRRC32*	A	G	0.945	0.892	1.001	5.33×10^{-2}	25424	0.863	0.820	0.908	1.37×10^{-8}	33086	1.52×10^{-8}	2.12×10^{-2}	0.898	0.864	0.933	3.09×10^{-8}	58510
rs7171171	15	38907041	*RASGRP1*	G	A	1.029	0.981	1.079	2.45×10^{-1}	25431	1.136	1.090	1.184	1.41×10^{-9}	33105	5.45×10^{-9}	2.03×10^{-3}	1.089	1.055	1.123	9.48×10^{-8}	58536
rs2368473	17	32534215	*CCL2*	G	A	1.038	0.998	1.079	6.31×10^{-2}	25431	1.103	1.067	1.141	1.29×10^{-8}	33101	1.64×10^{-8}	1.94×10^{-2}	1.074	1.047	1.102	3.64×10^{-8}	58532
rs72832915	17	37834808	*PGAP3*	A	G	1.034	0.992	1.077	1.13×10^{-1}	25429	1.110	1.071	1.150	1.07×10^{-8}	33103	2.19×10^{-8}	1.04×10^{-2}	1.076	1.048	1.106	8.63×10^{-8}	58532
UC autosomes																						
rs2523971	6	29938258	*HLA-A; HCG9*	A	C	0.965	0.924	1.008	1.11×10^{-1}	24304	0.874	0.836	0.913	1.30×10^{-9}	29540	2.78×10^{-9}	1.64×10^{-3}	0.918	0.890	0.947	5.79×10^{-8}	53844
rs2269426	6	32076499	*TNXB*	A	G	0.872	0.837	0.908	3.55×10^{-11}	24302	0.971	0.934	1.010	1.46×10^{-1}	29537	1.01×10^{-10}	1.80×10^{-4}	0.922	0.896	0.948	1.59×10^{-8}	53839
rs2239805	6	32411376	*HLA-DRA*	C	A	0.953	0.904	1.005	7.49×10^{-2}	24302	0.839	0.796	0.885	1.26×10^{-10}	29538	2.05×10^{-10}	9.00×10^{-4}	0.895	0.862	0.929	7.03×10^{-9}	53840
rs72834724	10	64441480	*ZNF365; ADO*	G	A	1.345	1.219	1.484	3.73×10^{-9}	24301	1.025	0.931	1.130	6.12×10^{-1}	29526	2.42×10^{-8}	1.22×10^{-4}	1.172	1.094	1.256	6.75×10^{-6}	53827
rs62037369	16	28883841	*SH2B1*	A	G	1.041	1.000	1.084	5.23×10^{-2}	24307	1.120	1.077	1.165	1.64×10^{-8}	29536	1.77×10^{-8}	1.10×10^{-2}	1.081	1.051	1.112	6.59×10^{-8}	53843
IBD autosomes																						
rs35812145	2	61391300	*C2orf74*	G	A	1.036	0.996	1.076	7.54×10^{-2}	33199	1.134	1.095	1.175	1.67×10^{-12}	40493	2.98×10^{-12}	5.94×10^{-4}	1.089	1.061	1.117	1.37×10^{-10}	73692
rs1990760	2	163124051	*IFIH1*	G	A	1.032	0.999	1.066	5.56×10^{-2}	33197	1.095	1.063	1.127	9.96×10^{-10}	40498	1.22×10^{-9}	7.89×10^{-3}	1.066	1.044	1.090	5.68×10^{-9}	73695
rs116630553	5	131949174	*RAD50*	A	G	0.917	0.838	1.003	5.92×10^{-2}	33200	0.791	0.728	0.860	3.53×10^{-8}	40502	4.14×10^{-8}	1.80×10^{-2}	0.847	0.797	0.900	1.01×10^{-7}	73702
rs11754821	6	31077614	*HCG22; C6orf15*	C	G	1.043	0.991	1.099	1.08×10^{-1}	33197	1.156	1.103	1.212	1.74×10^{-9}	40498	3.60×10^{-9}	4.24×10^{-3}	1.104	1.066	1.143	3.08×10^{-8}	73695
rs1265061	6	31078257	*C6orf15*	A	C	0.978	0.946	1.011	1.87×10^{-1}	33197	0.911	0.884	0.939	1.15×10^{-9}	40498	3.66×10^{-9}	1.95×10^{-3}	0.940	0.920	0.962	6.51×10^{-8}	73695
rs3128987	6	31434198	*HCP5; HCG26*	G	A	0.975	0.936	1.016	2.27×10^{-1}	33200	0.881	0.849	0.915	5.25×10^{-11}	40502	2.06×10^{-10}	3.74×10^{-4}	0.923	0.898	0.949	1.63×10^{-8}	73702
rs1150755	6	32038550	*TNXB*	A	G	0.960	0.918	1.005	8.00×10^{-2}	33200	0.854	0.819	0.891	2.12×10^{-13}	40502	4.11×10^{-13}	1.97×10^{-4}	0.902	0.875	0.930	5.21×10^{-11}	73702
rs2395158	6	32374595	*BTNL2*	G	A	0.963	0.921	1.007	9.49×10^{-2}	33199	0.879	0.844	0.915	4.76×10^{-10}	40502	9.16×10^{-10}	2.84×10^{-3}	0.916	0.889	0.944	1.10×10^{-8}	73701
rs4917014	7	50305863	*IKZF1*	C	A	1.033	0.999	1.068	6.09×10^{-2}	33200	1.091	1.059	1.124	1.53×10^{-8}	40498	1.88×10^{-8}	1.68×10^{-2}	1.065	1.041	1.089	4.75×10^{-8}	73698
rs61839660	10	6094697	*IL2RA*	A	G	1.027	0.973	1.084	3.28×10^{-1}	33198	1.152	1.097	1.209	1.03×10^{-8}	40496	4.61×10^{-8}	1.99×10^{-3}	1.095	1.056	1.135	8.76×10^{-7}	73694
rs76458677	10	35548225	*CCNY*	G	A	1.027	0.969	1.089	3.72×10^{-1}	33190	1.173	1.112	1.238	5.86×10^{-9}	40492	2.87×10^{-8}	1.01×10^{-3}	1.104	1.061	1.148	1.02×10^{-6}	73682
rs12268645	10	101328018	*NKX2-3;SLC25A28*	A	C	0.970	0.936	1.004	8.03×10^{-2}	33200	0.911	0.882	0.940	5.65×10^{-9}	40502	8.96×10^{-9}	8.69×10^{-3}	0.937	0.915	0.959	4.05×10^{-8}	73702

Abbreviations: CD, Crohn's disease; CI, confidence interval; IBD, inflammatory bowel disease; OR, odds ratio; SNP, single nucleotide polymorphism; UC, ulcerative colitis.
[a]For each disease subtype analyses, SNPs in linkage disequilibrium $r^2 < 0.5$.

associations further highlighting potential sex-dimorphism within MHC (Supplementary Table 3 and Supplementary Figure 1). Moreover, we examined whether any variants revealed sex-dimorphic genome-wide significance with sex-specific patterns of association for clinical variables. rs9673419/*NKD1* ($P_{2df\text{-}sex\text{-}dimorphic} = 1.7 \times 10^{-8}$, $P_{Female} = 1.5 \times 10^{-8}$) and rs621701/*C2* ($P_{2df\text{-}sex\text{-}dimorphic} = 2.9 \times 10^{-9}$, $P_{Male} = 2.3 \times 10^{-9}$) demonstrated association with colonic-only CD and age at diagnosis, respectively (**Supplementary Table 2B**). For CD inflammatory disease behavior, rs35419456/*THADA* exhibited genome-wide significance with the 1-df sex heterogeneity test ($P_{1df\text{-}sex\text{-}heterogeneity} = 1.0 \times 10^{-8}$) with comparable significance yet opposite directions of effect between sexes ($P_{2df\text{-}sex\text{-}dimorphic} = 6.9 \times 10^{-8}$; $P_{Sex\text{-}combined} = .78$) (**Supplementary Table 2B**).

We also examined SNPs that were associated with IBD in the largest study published to date.[15] Of the reported variants or proxies ($r^2 \geq 0.8$) available in our dataset, only 3 SNPs demonstrated association at $P < 5 \times 10^{-7}$ for one sex only, including rs61839660/*IL2RA* ($P_{Female} = 1.0 \times 10^{-8}$), rs16967103/*RASGRP1* ($P_{Female} = 8.1 \times 10^{-8}$), and rs254560/*C2orf66* ($P_{Female} = 3.6 \times 10^{-7}$), with only rs61839660 demonstrating genome-wide significance in 2-df sex-dimorphic meta-analyses (**Supplementary Table 4**). These results further emphasize the importance of exploring the interplay of sex and genetics in IBD beyond what may be identified with conventional analyses.

IBD-associations with chrX variants

We analyzed chrX SNPs by sex-stratified regression analyses making no assumptions of XCI status, as well as under assumptions of complete and uniform XCI, in addition to XCI-Escape. **Table 4** shows results for SNPs with $P < 4 \times 10^{-5}$ in CD. Given the 1262 chrX variants available, we considered a Bonferroni-corrected threshold of 4×10^{-5} for each analysis and a threshold of 5×10^{-6} to account for multiple tests performed. Differences in the observed level of significance for a variant tested under a model of XCI vs XCI-Escape may be reflective of the status of inactivation for a given locus in which a variant resides. rs2230488 in *RPS6KA3* exhibited the most robust significance observed under a model of XCI-Escape, while consistently demonstrating association at $P < 5 \times 10^{-6}$ across all analyses ($P_{XCI\text{-}Complete} = 1.1 \times 10^{-7}$ vs $P_{XCI\text{-}Escape} = 4.2 \times 10^{-8}$; $P_{Fisher\text{-}combined} = 7.5 \times 10^{-7}$, $P_{Stouffer\text{-}combined} = 8.5 \times 10^{-7}$). Similarly, rs34519770 in *MAP7D2* ($r^2 = 0.66$ with rs2230488/*RPS6KA3*) was more strongly associated under a model of XCI-Escape ($P = 2.4 \times 10^{-6}$), as well as with weighted Stouffer's combined ($P = 4.1 \times 10^{-6}$). rs1279816 (*STAG2*/*SH2D1A* intergenic), however, exhibited the greatest significance under assumptions of complete XCI ($P = 2.1 \times 10^{-5}$) yet did not reach $P < 5 \times 10^{-6}$ across any analyses (**Table 4**). No SNPs were associated with UC or IBD at $P < 4 \times 10^{-5}$. Our chrX associations with CD stress the importance of investigating the neglected chromosome X in IBD susceptibility and highlight candidate IBD loci for further functional studies.

Discussion

The "gender gap" is well-established, with a greater prevalence in females for many (systemic lupus erythematosus, rheumatoid arthritis, Sjögren's syndrome, multiple sclerosis) but not all (ankylosing spondylitis) immune-mediated inflammatory diseases.[1] There are numerous factors underlying this observed sexual dimorphism in both prevalence and severity, including sex differences in the immune system and immune response, hormones, genetics, and environmental elements. There has been a pronounced shift in recent years toward considering sex as a critical variable in research. For IBD, while sex differences in clinical observations have been documented, a thorough investigation of sex differences in genetic association has been lacking. Here, we present the first and largest systematic investigation of the role for sex in IBD genetics.

As with other immune-mediated diseases, sex differences in IBD have been reported for natural history and disease complications, including with age of onset, disease location, presence of extraintestinal manifestations, and risk for surgery.[3-5] In our study of the largest dataset to date, we identified significant phenotypic differences between the sexes for disease location, extent, and complications. We observed males to more likely have both upper GI disease and perianal disease complications in CD. Males were less likely to have both colonic-only disease location in CD and also more extensive disease in UC. We did not observe any significant difference in disease behavior between sexes (**Table 1**). Greater female predominance has been previously reported for CD, yet the interpretation of this is complex, given the effects of estrogen signaling on the immune system and epithelial homeostasis.[2,6,13] In vivo studies support a critical role of estrogen signaling in sex-based differences in IBD.[26,27] Male SAMP1/YitFc mice have been shown to expand their intestinal T-reg cell population and reduce inflammation in response to exogenous estrogen, while female SAMP1/YitFc mice are resistant to the effects of exogenous estrogen, exhibiting dysregulation of estrogen receptor β–specific signaling, impaired T-reg suppressive functions, and more severe ileitis.[27] Furthermore, female patients with active CD have been shown to demonstrate reduced estrogen receptor β expression in intestinal mucosal and peripheral T cells and aberrant T-reg–specific expression of transcription factor GILZ, resulting in a loss of functional T-reg suppression.[26] Such preclinical studies warrant the importance of thoroughly investigating sex differences in IBD.

There have been a limited number of studies investigating sex-specific genetic associations in IBD. R30Q in *DLG5* has previously been shown to be a sex-specific susceptibility variant conferring risk of CD in males.[11] *DLG5* R30Q was not significant in our dataset. Female-specific associations have also been reported for *IL10*, *IL23R*, and *ATG16L1*, but most findings have not been replicated.[9,10,28] We did not observe sex-specific associations in these loci. Genome-wide significance was observed in both sexes for rs6752107/*ATG16L1* and rs3024505/*IL10*, and we observed a more robust female sex-stratified signal, in keeping with previous reports. The opposite trend was observed for rs11209026/*IL23R* (**Supplementary Table 4**). Moreover, most previously reported IBD-associated variants[15] did not demonstrate any sex-specific patterns (**Supplementary Table 4**).

The aim of our study was to comprehensively examine whether sex plays a significant role in the genetic etiology of IBD susceptibility. For this study, we utilized a 2-df sex-dimorphic test that has been shown to be comparable to sex-combined analyses in the absence of heterogeneity between sexes, yet more powerful in the presence of potential heterogeneity. The 2-df sex-dimorphic test also allows for

Table 4: Chromosome X associations with CD.

Nearest gene(s)	SNP	Position (hg19)	Allele	Cohort		Sex-stratified case-control					XCI-Complete		XCI-Escape	
						P	OR	95% CI	*P* (Fisher combined)	*P* (Stouffer combined)=	*P*	OR	*P*	OR
RPS6KA3	rs2230488	20,204,461	A	Cedars	Male	.22	1.06	0.900-1.244	5.01×10^{-3}	2.73×10^{-3}	6.21×10^{-3}	1.11	1.80×10^{-3}	1.18
					Female	2.72×10^{-3}	1.21	1.08-1.362						
				IIBDGC	Male	3.13×10^{-3}	1.06	0.984-1.140	1.01×10^{-5}	2.19×10^{-6}	3.82×10^{-6}	1.07	2.95×10^{-6}	1.10
					Female	2.12×10^{-4}	1.09	1.043-1.143						
				Combined	–	–	–	–	7.52×10^{-7}[a]	8.52×10^{-7}[a]	1.10×10^{-7}[a]	1.08[a]	4.22×10^{-8}[a]	1.11[a]
MAP7D2	rs34519770	20,071,046	G	Cedars	Male	.41	1.03	0.896-1.192	.14	.07	.09	1.06	.06	1.09
					Female	.08	1.10	0.996-1.224						
				IIBDGC	Male	.05	1.03	0.969-1.102	5.49×10^{-5}	2.00×10^{-5}	7.31×10^{-5}	1.05	1.44×10^{-5}	1.08
					Female	8.12×10^{-5}	1.08	1.041-1.127						
				Combined	–	–	–	–	1.11×10^{-4}[a]	4.08×10^{-6}[a]	1.66×10^{-5}[a]	1.05[a]	2.38×10^{-6}[a]	1.08[a]
STAG2, SH2D1A	rs1279816	123,355,614	A	Cedars	Male	2.30×10^{-2}	0.91	0.781-1.058	3.01×10^{-2}	1.25×10^{-2}	9.63×10^{-3}	0.92	2.09×10^{-2}	0.90
					Female	.21	0.93	0.836-1.033						
				IIBDGC	Male	2.54×10^{-2}	0.96	0.903-1.027	1.26×10^{-3}	3.35×10^{-4}	3.63×10^{-4}	0.96	6.40×10^{-4}	0.94
					Female	4.98×10^{-3}	0.95	0.909-0.982						
				Combined	–	–	–	–	3.80×10^{-4}[a]	2.34×10^{-4}[a]	2.11×10^{-5}[a]	0.95[a]	6.41×10^{-5}[a]	0.94[a]

Chromosome X variants were tested for association with CD under models of complete and uniform XCI (XCI-Complete) or escape from XCI (XCI-Escape) in females, and compared with sex-stratified case-control analyses, combined using weighted Fisher's or Stouffer's methods, in Cedars-Sinai Medical Center (Cedars) (male cases n = 1420 and controls n = 2621; female cases n = 1262 and controls n = 2874) and IIBDGC (male cases n = 7214 and controls n = 14 074; female cases n = 9493 and controls n = 19 452).
Abbreviations: CD, Crohn's disease; CI, confidence interval; IIBDGC, International IBD Genetics Consortium; OR, odds ratio; SNP, single nucleotide polymorphism.

utilization of the full sample size as compared with sex-specific analyses.[17] We identified 3 novel loci at $P_{2df\text{-}sex\text{-}dimorphic} < 5 \times 10^{-8}$ including chr9q22, *CARMIL1*, and *UBASH3A* (**Table 2**). We observed association with IBD and 2 variants in *CARMIL1* (aka *LRRC16A*), encoding a leucine-rich repeat–containing membrane-associated protein detected in a wide variety of tissues, including the terminal ileum. *CARMIL1* has recently been shown to associate with interleukin (IL)-1 signaling molecules and play a role in IL-1–induced ERK activation.[29] This locus has been associated in genetic studies of numerous traits, with 1 of our 2 IBD-associated variants (rs7752195) implicated in an autism-schizophrenia meta-analysis.[30] Our novel association with UC was with *UBASH3A*, a gene involved in regulation of T cell signaling and not previously identified in any recent large-scale IBD genetic studies,[14,15,31] although nominal association has been reported with UC.[32] We achieved genome-wide significance for rs1893592, an eQTL for *UBASH3A* in transverse colon, terminal ileum, and whole blood, among other tissues, which has also demonstrated association with multiple immune-mediated diseases such as primary sclerosing cholangitis and rheumatoid arthritis.[33,34] rs1893592 is located within a splice consensus sequence downstream of exon 10, with the C allele predicted to disrupt the 5′ splice donor sequence, potentially leading to nonstop-mediated messenger RNA decay.[34] Further investigation is needed to establish the functional consequences of this variant on UBASH3A protein and its implications in IBD.

Variants achieving $P_{2df\text{-}sex\text{-}dimorphic} < 5 \times 10^{-8}$ were next evaluated for sex-specific association ($P < 5 \times 10^{-7}$ in one sex only). Multiple ($r^2 < 0.35$) variants in chr2p15 demonstrated the most significant sex-dimorphic *P* values and were associated with a female-specific increased risk in CD and IBD (**Table 3**). This locus harboring genes *C2orf74*, *USP34*, and *PUS10* has previously been implicated in IBD and other chronic inflammatory diseases.[15,32] While no one specific gene in chr2p15 has been definitively implicated in IBD pathogenesis, variants reported here are eQTLs and splicing quantitative trait loci for *C2orf74* in numerous tissues, including terminal ileum and sigmoid and transverse colon, and reside in a region demonstrating histone modifications consistent with strong transcription in colonic mucosa and the sigmoid colon, among others. In addition, one-third of our observed sex-dimorphic associations resided within the MHC (**Table 3**). While countless IBD studies have documented multiple independent genetic associations in both HLA and non-HLA genes within the MHC, a recent transethnic analysis of HLA in UC did not reveal evidence of sex-biased association using a case-only analytic approach.[35] Sex differences are well established in the context of both innate and adaptive immunity.[6] Coupled with the known role of MHC in IBD, our sex-dimorphic observations within this locus were not unexpected, and our results warrant functional studies of the role of sex and MHC associations with IBD.

We also identified association with phospholipase *PGAP3* (**Table 3**). rs72832915 exhibited female-specific sex-dimorphic association with CD and resides in a gene-dense region with evidence of association with IBD and sex-biased autoimmune diseases including rheumatoid arthritis.[15,33] While rs72832915 is intronic to *PGAP3*, this variant falls within an open chromatin binding site for HNF4A, a transcription factor with a role in intestinal epithelial homeostasis, and is an eQTL for *GSDMA* in the transverse colon, among others. Furthermore a recent study evaluating the effect of sex on gene expression across tissues identified *GSDMA*, a member of the gasdermin family of pore-forming proteins that can interfere with cell membrane integrity and trigger inflammatory pyroptosis, as among the top 500 sex-biased genes in the transverse colon.[8,36] Our results warrant a consideration of sex for future studies of *GSDMA* in IBD and colitis to better understand the role of sex underlying the association at these loci.

Additionally, we identified potential novel loci highlighted by variants exhibiting nominal significance with 2-df sex-dimorphic meta-analysis ($5 \times 10^{-8} < P < 5 \times 10^{-7}$), including rs57275892/*PSMA6* with CD and rs980263457/*OS9* with UC (**Supplementary Table 3**). *PSMA6* encodes a component of the 20S core proteasome complex involved in the proteolytic degradation of most intracellular proteins. While the most recent, largest European IBD genetic association study has not implicated this locus, association with rs57275892 has been reported in an East Asian IBD study.[20] Of particular interest to our sex-dimorphic association observed at this locus is evidence of sex-biased expression for PSMA6 in both the sigmoid colon and terminal ileum.[8] In addition, differential gene expression analyses from our previously published transcriptomic data of uninvolved small bowel CD surgical subjects[37] highlighted decreased expression of PSMA6 in females (log fold change = -0.31 and -0.44; $P_{Adjusted} = 3.1 \times 10^{-7}$ and 5.8×10^{-7}), further supportive of a potential role for sex at this locus. The second potential novel locus is *OS9*, which encodes an endoplasmic reticulum (ER) lectin involved in ER-associated protein degradation that protects intestinal epithelial barrier function under hypoxic conditions in vitro.[38] Our observed associations with UC and *OS9* as well as with rs6088728 in *EDEM2*, encoding another ER lectin, coupled with a recognized role for ER stress in intestinal epithelial barrier function in the context of IBD,[39] suggest a promising avenue for further investigation of the role of sex underlying ER stress pathways in IBD pathogenesis.

ChrX associations have thus far regularly been overlooked due to a lack of robust statistical methodology. It has been reported that up to a quarter of X-linked genes demonstrate complete or variable XCI-Escape in healthy females,[12] which may result in sex-biased gene expression across human tissues.[8] With minimal literature on chrX genetic association in IBD, we comprehensively evaluated associations with chrX variants, including under models of complete and uniform XCI and XCI-Escape, to identify candidate IBD X-linked loci. Of the top 2 novel exonic associations ($r^2 = 0.66$) in *RPS6KA3* and *MAP7D2* on Xp22.12 (**Table 4**), the lead SNP rs2230488 lies within a region of histone modifications marks consistent with strong transcription in various tissues including $CD4^+$ T helper cells, rectal mucosa, the sigmoid colon, and the small intestine. Both variants demonstrate evidence of eQTLs for nearby genes *EIF1AX* and *Cxorf23* in adipose, colon, thyroid, and whole blood. The eukaryotic translation initiation factor *EIF1AX* is located within a chrX stratum with a high chance of escaping XCI and displays higher expression in females for several tissues including the colon and terminal ileum.[8,40] While we did not consider XCI patterns across different tissues and deeper investigations of sex-specific expression patterns are warranted, these *EIF1AX* data are consistent with our observations of rs2230488 and rs34519770 having strongest associations under models of XCI-Escape. Differential expression between sexes for

MAP7D2 and *EIF1AX* has also recently been demonstrated in whole blood.[7,8] Our analyses identify association at Xp22.12 and highlight a potential role for *EIF1AX*, a gene not previously implicated in IBD or other immune-mediated diseases.

Our study highlights important findings, but some limitations should be acknowledged. While we identified genome-wide significance with the 2-df sex-dimorphic test, with the exception of rs35419456/*THADA* and CD disease behavior (**Supplementary Table 2B**), no other variants reached genome-wide significance with the 1-df Cochran's Q test for sex heterogeneity. However, this is in keeping with previous reports that sex heterogeneity test is generally underpowered in comparison with the 2-df sex-dimorphic test, when effects in the same direction are larger in or are specific to one sex.[17,24] We also recognize the difference in sample size between sexes, which may impact power in our sex-specific analyses, yet one benefit of implementing the 2-df sex-dimorphic meta-analytic approach is the ability to utilize the full dataset to include both sexes in analyses. There is increasing evidence of sex-based differences in IBD epidemiology in non-European cohorts.[41] In this study, we were limited to a European ancestry population as well as polymorphisms present on the Immunochip array. Follow-up investigations would greatly benefit from cross-ancestry analyses in addition to more comprehensive whole-genome platforms, particularly for chrX coverage and expanding MHC studies to include HLA alleles. Furthermore, these results would benefit from independent replication, particularly for variants at potentially novel loci. Finally, while we investigated chrX associations under models of XCI and XCI-Escape, we did not evaluate XCI patterns, which are known to be quite variable and tissue specific, across the different IBD-relevant tissues. In addition, while one benefit of consortia-based studies is a dramatic increase in power, this often comes at the expense of missing subphenotypic and demographic characteristics, as observed particularly with our age-related variables. We also cannot rule out that the associations reported here are not due to thus far unidentified IBD risk factors that may vary by sex.

In summary, we present here the largest study to date investigating IBD phenotypes across the sexes and observed significant differences for CD location, perianal disease, and UC extent. We also demonstrated that the 2-df sex-dimorphic and chrX analyses have the potential to uncover genetic factors of sexual dimorphism that may contribute to refining our understanding of IBD risk. Our analyses identified 3 novel loci previously unreported in recent large-scale IBD GWASs. We also characterized sex-dimorphic associations at *C2orf74* and within the MHC, in addition to *PGAP3/GSDMA*, *PSMA6*, and *OS9*. Given the numerous immune-related genes located on chrX, we identified a potentially novel locus on chromosome Xp22 (*RPS6KA3/MAP7D2/EIF1AX*). Studies such as ours help us move toward elucidating role of sex-dependent and chrX variation in IBD genetic architecture and provide important clues about disease etiology.

Supplementary data

Supplementary data is available at *Inflammatory Bowel Diseases* online.

Acknowledgments

We are thankful to all clinicians, researchers, coordinators, and patients who have contributed time, data, and samples to the MIRIAD Biobank as well as the National Institute of Diabetes and Digestive and Kidney Diseases and International Inflammatory Bowel Disease Genetics Consortia. Consortia membership information is provided in the Supplementary Data. Written informed consent was obtained from all Cedars-Sinai Medical Center subjects, and study protocol and data collection were approved by Cedars-Sinai Medical Center Institutional Review Board. For the International IBD Genetics Consortium, all patients provided written informed consent approved by the Institutional Review Board of each recruiting site.

Funding

This work was supported by internal funds from the F. Widjaja Foundation Inflammatory Bowel Disease Institute. The Cedars-Sinai MIRIAD IBD Biobank is supported by the F. Widjaja Foundation Inflammatory Bowel Disease Institute, National Institutes of Health/National Institute of Diabetes and Digestive and Kidney Diseases grants (P01 DK046763 and U01 DK062413), and the Leona M and Harry B Helmsley Charitable Trust (DPBM).

Conflicts of Interest

Cedars-Sinai has financial interests in Prometheus Biosciences, Inc, a company that has access to the data and specimens in Cedars-Sinai's MIRIAD Biobank (including some of the genetic data used in this study) and seeks to develop commercial products. D.P.B.M. owns stock in Prometheus Biosciences Inc. A.A.P., D.L., and D.P.B.M. are consultants for Prometheus Biosciences, Inc. D.P.B.M. has served as a consultant for Gilead, Pfizer, Boehringer Ingelheim, Qu Biologics, Bridge Biotherapeutics, Takeda, and Palatin Technologies; and received grant support from Janssen. The other authors disclose no conflicts.

Data Availability

Data underlying this article are available in the article and in its online supplementary material. Summary statistics will be made available through the National Institute of Diabetes and Digestive and Kidney Diseases IBD Genetics Consortium (IBDGC) Data Commons portal currently under development. The IBDGC Data Commons supports the management, analysis, and sharing of genetic data to support the vision and mission of the IBDGC.

References

1. Ngo ST, Steyn FJ, McCombe PA. Gender differences in autoimmune disease. *Front Neuroendocrinol*. 2014;35(3):347-369.
2. Betteridge JD, Armbruster SP, Maydonovitch C, Veerappan GR. Inflammatory bowel disease prevalence by age, gender, race, and geographic location in the U.S. military health care population. *Inflamm Bowel Dis*. 2013;19(7):1421-1427.
3. Shah SC, Khalili H, Gower-Rousseau C, et al. Sex-based differences in incidence of inflammatory bowel diseases-pooled analysis of population-based studies from western countries. *Gastroenterology*. 2018;155(4):1079-1089.e3.

4. Severs M, Spekhorst LM, Mangen MJ, et al. Sex-related differences in patients with inflammatory bowel disease: results of 2 prospective cohort studies. *Inflamm Bowel Dis*. 2018;24(6):1298-1306.
5. Greuter T, Manser C, Pittet V, et al. Gender differences in inflammatory bowel disease. *Digestion*. 2020;101(suppl 1):98-104.
6. Klein SL, Flanagan KL. Sex differences in immune responses. *Nat Rev Immunol*. 2016;16(10):626-638.
7. Lopes-Ramos CM, Chen CY, Kuijjer ML, et al. Sex differences in gene expression and regulatory networks across 29 human tissues. *Cell Rep*. 2020;31(12):107795.
8. Oliva M, Munoz-Aguirre M, Kim-Hellmuth S, et al. The impact of sex on gene expression across human tissues. *Science*. 2020;369(6509):eaba3066.
9. Tedde A, Laura Putignano A, Bagnoli S, et al. Interleukin-10 promoter polymorphisms influence susceptibility to ulcerative colitis in a gender-specific manner. *Scand J Gastroenterol*. 2008;43(6):712-718.
10. Liu LY, Schaub MA, Sirota M, Butte AJ. Transmission distortion in Crohn's disease risk gene ATG16L1 leads to sex difference in disease association. *Inflamm Bowel Dis*. 2012;18(2):312-322.
11. Friedrichs F, Brescianini S, Annese V, et al. Evidence of transmission ratio distortion of DLG5 R30Q variant in general and implication of an association with Crohn disease in men. *Hum Genet*. 2006;119(3):305-311.
12. Tukiainen T, Villani AC, Yen A, et al.; GTEx Consortium. Landscape of X chromosome inactivation across human tissues. *Nature*. 2017;550(7675):244-248.
13. Fish EN. The X-files in immunity: sex-based differences predispose immune responses. *Nat Rev Immunol*. 2008;8(9):737-744.
14. Jostins L, Ripke S, Weersma RK, et al.; International IBD Genetics Consortium (IIBDGC). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491(7422):119-124.
15. de Lange KM, Moutsianas L, Lee JC, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet*. 2017;49(2):256-261.
16. Gao F, Chang D, Biddanda A, et al. XWAS: a software toolset for genetic data analysis and association studies of the X chromosome. *J Hered*. 2015;106(5):666-671.
17. Magi R, Lindgren CM, Morris AP. Meta-analysis of sex-specific genome-wide association studies. *Genet Epidemiol*. 2010;34(8):846-853.
18. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*. 2015;4(1):7.
19. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015;31(21):3555-3557.
20. Hong M, Ye BD, Yang SK, et al. Immunochip meta-analysis of inflammatory bowel disease identifies three novel loci and four novel associations in previously reported loci. *J Crohns Colitis*. 2018;12(6):730-741.
21. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22(9):1790-1797.
22. GTEx C. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585.
23. Vosa U, Claringbould A, Westra HJ, et al.; BIOS Consortium. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet*. 2021;53(9):1300-1310.
24. Lagou V, Magi R, Hottenga JJ, et al.; Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC). Sex-dimorphic genetic effects and novel loci for fasting glucose and insulin variability. *Nat Commun*. 2021;12(1):24.
25. Goyette P, Boucher G, Mallon D, et al.; International Inflammatory Bowel Disease Genetics Consortium. High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nat Genet*. 2015;47(2):172-179.
26. Goodman WA, Bedoyan SM, Havran HL, Richardson B, Cameron MJ, Pizarro TT. Impaired estrogen signaling underlies regulatory T cell loss-of-function in the chronically inflamed intestine. *Proc Natl Acad Sci U S A*. 2020;117(29):17166-17176.
27. Goodman WA, Garg RR, Reuter BK, Mattioli B, Rissman EF, Pizarro TT. Loss of estrogen-mediated immunoprotection underlies female gender bias in experimental Crohn's-like ileitis. *Mucosal Immunol*. 2014;7(5):1255-1265.
28. Lin Z, Poritz L, Franke A, et al. Genetic association of nonsynonymous variants of the IL23R with familial and sporadic inflammatory bowel disease in women. *Dig Dis Sci*. 2010;55(3):739-746.
29. Wang Q, Notay K, Downey GP, McCulloch CA. The leucine-rich repeat region of CARMIL1 regulates IL-1-mediated ERK activation, MMP expression, and collagen degradation. *Cell Rep*. 2020;31(13):107781.
30. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16 000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol Autism*. 2017;8:21.
31. Ge Y, Paisie TK, Chen S, Concannon P. UBASH3A regulates the synthesis and dynamics of TCR-CD3 complexes. *J Immunol*. 2019;203(11):2827-2836.
32. Ellinghaus D, Jostins L, Spain SL, et al.; International IBD Genetics Consortium (IIBDGC). Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet*. 2016;48(5):510-518.
33. Okada Y, Wu D, Trynka G, et al.; RACI consortium. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*. 2014;506(7488):376-381.
34. Ji SG, Juran BD, Mucha S, et al.; UK-PSC Consortium. Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. *Nat Genet*. 2017;49(2):269-273.
35. Degenhardt F, Mayr G, Wendorff M, et al.; MAAIS Recruitment Center. Transethnic analysis of the human leukocyte antigen region for ulcerative colitis reveals not only shared but also ethnicity-specific disease associations. *Hum Mol Genet*. 2021;30(5):356-369.
36. Liu X, Xia S, Zhang Z, Wu H, Lieberman J. Channelling inflammation: gasdermins in physiology and disease. *Nat Rev Drug Discov*. 2021;20(5):384-405.
37. Potdar AA, Li D, Haritunians T, et al. Ileal gene expression data from Crohn's disease small bowel resections indicate distinct clinical subgroups. *J Crohns Colitis*. 2019;13(8):1055-1066.
38. Sun L, Xu C, Chen G, et al. A novel role of OS-9 in the maintenance of intestinal barrier function from hypoxia-induced injury via p38-dependent pathway. *Int J Biol Sci*. 2015;11(6):664-671.
39. Graham DB, Xavier RJ. Pathway paradigms revealed from the genetics of inflammatory bowel disease. *Nature*. 2020;578(7796):527-539.
40. Schurz H, Salie M, Tromp G, Hoal EG, Kinnear CJ, Möller M. The X chromosome and sex-specific effects in infectious disease susceptibility. *Hum Genomics*. 2019;13(1):2.
41. Shah SC, Khalili H, Chen CY, et al. Sex-based differences in the incidence of inflammatory bowel diseases-pooled analysis of population-based studies from the Asia-Pacific region. *Aliment Pharmacol Ther*. 2019;49(7):904-911.